

---

**UNIVERSIDAD AUTONOMA DE MADRID**

**ESCUELA POLITECNICA SUPERIOR**



**Grado en Ingeniería Informática**

## **TRABAJO FIN DE GRADO**

**HERRAMIENTA INFORMÁTICA DE APOYO AL USUARIO  
EN ENTORNOS EDUCATIVOS EN LÍNEA**

**Alejandro Rodríguez Ranera**  
**Tutor: Ruth Cobos Pérez**

**Junio 2018**



# **HERRAMIENTA INFORMÁTICA DE APOYO AL USUARIO EN ENTORNOS EDUCATIVOS EN LÍNEA**

**AUTOR: Alejandro Rodríguez Ranera**

**TUTOR: Ruth Cobos Pérez**

**Dpto. de Ingeniería Informática  
Escuela Politécnica Superior  
Universidad Autónoma de Madrid  
Junio de 2018**



# Resumen

Las nuevas tecnologías han ayudado a la generación, distribución y acceso al conocimiento. Esto ha permitido enriquecer y transformar la educación. Un claro ejemplo son los *Massive Open Online Courses* (MOOC), que permiten el acceso a miles de cursos y a su divulgación a través de internet. A pesar de las múltiples ventajas que ofrece la educación a distancia, también plantea nuevos retos a lo que enfrentarse, como la alta tasa de abandono y fracaso que presentan este tipo de cursos.

El objetivo de este Trabajo de Fin de Grado consiste en la creación de una herramienta (*edX-WS: Warning System for edX MOOC*) para la detección preventiva de usuarios con riesgo de no finalizar satisfactoriamente un curso en base a sus indicadores. Estos indicadores recolectan la actividad y las tareas realizadas por los estudiantes a lo largo de su aprendizaje. Además de detectar a estos usuarios, permite generar sugerencias y realizar exportaciones de datos comparativos para poder ayudar y seguir la evolución de cada estudiante. De esta manera, se pretende apoyar a los usuarios que se inscriben en cursos MOOC, para que consigan finalizar el curso y aprobarlo para poder obtener el certificado.

Este proyecto, se basa en los datos recopilados de los MOOC de la UAM y hace uso de la información generada por la herramienta *edX-MAS+ (edX-MAS+: Model Analyzer System for edX MOOC)* para poder generar y entrenar modelos predictivos y de esta forma permitir descubrir usuarios con un alto riesgo de no finalizar un curso o de suspenderlo (y, por tanto, no obtener el certificado). Además, posibilita ver comparativas diarias de los indicadores para cada uno de ellos, con usuarios que si obtuvieron el certificado en años anteriores, comparativas grupales, descarga de ficheros *csv* con toda la información recopilada, exportación a ficheros *xlsx* de los datos utilizados para la generación de gráficas y la generación de sugerencias en base a las carencias detectadas en los indicadores de cada usuario.

La finalidad de esta aplicación es proporcionar una herramienta para que los gestores de los cursos MOOC de la UAM puedan analizar y realizar un seguimiento de los estudiantes inscritos a sus cursos. Pudiendo de esta manera ofrecer ayudas y lograr que posibles estudiantes en riesgo acaben finalizando el curso y obteniendo el certificado.

# Abstract

New technologies have helped to generation, distribution and access to knowledge. This has allowed enriching and transforming the education. A clear example is the *Massive Open Online Courses* (MOOC), which allows access to thousands of courses and their dissemination through the internet. Despite the many advantages offered by distance education, it also propose new challenges, such as the high dropout rate and the failure of this type of course.

The objective of this Final Degree Project is to create a tool (*edX-WS: Warning System for edX MOOC*) for the preventive detection of users at risk of not successfully completing a course based on their indicators. These indicators collect the activity and the tasks carried out by the students throughout their learning. In addition to detecting these users, it allows generating suggestions and exports comparative charts in order to help and follow the evolution of each student. In this way, it is intended to support users who enroll in MOOC courses, so that they can complete the course and approve it in order to obtain the certificate.

This project is based on the data collected from UAM MOOCs and makes use of the information generated by *edX-MAS+* tool (*edX-MAS+: Model Analyzer System for edX MOOC*) to generate and train predictive models and thus allow to discover users with a high risk of not finishing a course or of suspending it (and therefore, not obtaining the certificate). It also makes it possible to see daily comparisons of the indicators for each one of them with users who did obtain the certificate in previous years, group comparisons, download of *csv* files with all the information collected, export to *xlsx* files of the data used to generate charts and generation of suggestions to help users.

The purpose of this application is to provide a tool for UAM MOOCs managers to analyze and track students enrolled in their courses. This way, we can offer aid and manage to guide potential students at risk finish the course and obtain the certificate.

## **Palabras clave**

Modelos predictivos, Cursos Online Masivos y Abiertos, Análisis de Datos, Indicadores, Sugerencias, Recomendaciones, Sistema de Avisos, Exportación de Datos.

## **Keywords**

Predictive Models, Massive Open Online Courses, Data Analysis, Indicators, Suggestions, Recommendation, Warning System, Data Export.





## ***Agradecimientos***

Me gustaría agradecer enormemente a mi familia por hacerme llegar hasta aquí, por preocuparse por mí, por estar ahí siempre que ha hecho falta y por todo su cariño. Sin ellos nada de esto hubiera sido posible.

Dar las gracias a Carolina por enseñarme a sonreír hasta en los momentos más difíciles, por darme todo su cariño, por haberme hecho ser quien soy y por compartir junto a mí los días más felices de mi vida.

Gracias de todo corazón a Rone, Yunta, Iago, Lebron y Sanbón por ser lo mejor que me llevo de mi paso por la universidad y por ser a día de hoy mis mejores amigos.

También dar las gracias al resto de compañeros con los que he compartido mis días en la universidad, Ana, Diego, Rubio, Rober, Vadillo, Adri, Pedro y Óscar. Mencionar especialmente a Sandra porque sin ella mi etapa final de la carrera hubiese sido mucho más complicada.

Por supuesto agradecer a esos maravillosos compañeros de trabajo, Álvaro, Alberto, Miguel, Fran, Ana \* 2, Sergio, Aida y Mario. También agradecer a Víctor y Juanjo las facilidades que me han dado para poder terminar este proyecto y apoyarme hasta el final.

Por último agradecer a Ruth Cobos por darme la maravillosa oportunidad de realizar este proyecto y por ayudarme a finalizar mis estudios.



## INDICE DE CONTENIDOS

1	Introducción.....	1
1.1	Motivación.....	1
1.2	Objetivos.....	2
1.3	Organización de la memoria.....	3
2	Estado del arte .....	5
2.1	Educación a distancia .....	5
2.2	Importancia y potencial de los MOOC's.....	5
2.3	Predicción en cursos online .....	6
2.4	Test con modelos .....	7
2.5	Análisis de la herramienta edX-MAS+ y edX-WS .....	8
2.5.1	Leguajes de desarrollo y librerías.....	8
2.5.2	Arquitectura de edX-MAS+ .....	9
2.5.3	Indicadores disponibles .....	10
3	Diseño.....	12
3.1	Arquitectura Lógica.....	12
3.1.1	Capa de presentación.....	13
3.1.2	Capa de negocio.....	13
3.1.3	Capa de datos.....	13
3.2	Requisitos funcionales.....	14
3.3	Requisitos no funcionales.....	18
3.4	Diseño de los módulos.....	18
3.4.1	Módulo de predicción.....	19
3.4.2	Módulo de procesamiento .....	19
3.4.3	Módulo de visualización y exportación.....	19
4	Desarrollo .....	23
4.1	Estructura de los ficheros .....	23
4.1.1	Directorio /postgresqlstore .....	23
4.1.2	Directorio /R .....	23
4.1.3	Directorio /R/ws .....	24
4.2	Estructura de la Base de Datos .....	24
4.3	Implementación .....	26
4.3.1	ui.R .....	26
4.3.2	server.R.....	26
5	Integración, pruebas y resultados .....	29
5.1	Pruebas funcionales .....	29
5.1.1	Pruebas unitarias.....	29
5.1.2	Pruebas de integración.....	29
5.1.3	Pruebas de aceptación.....	30
5.2	Pruebas no funcionales .....	31
5.2.1	Pruebas de compatibilidad.....	31
6	Conclusiones y trabajo futuro.....	33
6.1	Conclusiones.....	33
6.2	Trabajo futuro .....	34
	Referencias .....	37
	Glosario .....	41
	Anexos.....	- 1 -
	A Manual de usuario.....	- 1 -

B Diagrama de Gantt .....- 8 -

## INDICE DE FIGURAS

FIGURA 2.5-1: LOGO R .....	8
FIGURA 2.5-2: LOGO POSTGRESQL .....	8
FIGURA 2.5-3: ARQUITECTURA EDX-MAS+ .....	9
FIGURA 3.1-1: ARQUITECTURA LÓGICA DE EDX-WS.....	12
FIGURA 3.4-1: MÓDULOS EDX-WS .....	18
FIGURA 4.2-1: TABLAS DE LA BASE DE DATOS.....	24
FIGURA 4.2-2: MODELO DE LA BASE DE DATOS .....	25
FIGURA 5.1-1: FICHEROS EXPORTADOS.....	31
FIGURA A-1: EXPLICACIÓN PANTALLA DE INICIO.....	- 1 -
FIGURA A-2: EXPLICACIÓN PANTALLA DE MODELOS PREDICTIVOS CARGADOS.....	- 2 -
FIGURA A-3: EXPLICACIÓN DE PARÁMETROS PARA GENERACIÓN DE MODELOS .....	- 2 -
FIGURA A-4: EXPLICACIÓN DE PARÁMETROS PARA GENERACIÓN DE MODELOS .....	- 3 -
FIGURA A-5: EXPLICACIÓN PANTALLA DE VISUALIZACIÓN DE LOS RESULTADOS DEL TEST.....	- 3 -
FIGURA A-6: EXPLICACIÓN DE LA SELECCIÓN DEL MODELO PARA REALIZAR EL TEST .....	- 4 -
FIGURA A-7: EXPLICACIÓN DE LA SELECCIÓN DEL CURSO PARA REALIZAR EL TEST .....	- 4 -
FIGURA A-8: EXPLICACIÓN DE LA SELECCIÓN DEL CURSO PARA REALIZAR EL TEST .....	- 5 -
FIGURA A-9: EXPLICACIÓN PARA CONSULTAR DETALLE Y EXPORTACIONES .....	- 5 -
FIGURA A-10: DESCARGA DE FICHEROS (EXPORTACIONES Y SUGERENCIAS) .....	- 6 -
FIGURA A-11: EXPLICACIÓN PANTALLA DE VISUALIZACIÓN DE DETALLE DE UN USUARIO .....	- 6 -
FIGURA A-12: EXPLICACIÓN PANTALLA DE VISUALIZACIÓN DE DETALLE DE GRUPOS .....	- 7 -
FIGURA B-1: DIAGRAMA DE GANTT 1.....	- 8 -
FIGURA B-2: DIAGRAMA DE GANTT 2.....	- 8 -

## INDICE DE TABLAS

TABLA 3-1: RF-01 .....	14
TABLA 3-2: RF-02 .....	15
TABLA 3-3: RF-03 .....	15
TABLA 3-4: RF-04 .....	16
TABLA 3-5: RF-05 .....	16
TABLA 3-6: RF-06 .....	17
TABLA 3-7: REQUISITOS NO FUNCIONALES .....	18
TABLA 4-1: ESTRUCTURA DE FICHEROS.....	23



# 1 Introducción

---

En este primer capítulo se expone la motivación que ha llevado a la realización de este Trabajo de Fin de Grado, los objetivos a cumplir y la organización de la memoria.

## 1.1 Motivación

Los MOOCs están cobrando gran importancia en la actualidad debido a que permiten llegar a un gran número de estudiantes. Facilitando la difusión de conocimientos mediante cursos gracias a internet. Hasta la fecha se ha transmitido información y noticias a distancia, por ejemplo, vía televisión, internet, radio, periódicos, etc... pero la educación presencial ha imperado sin necesidad de utilizar estas tecnologías. Actualmente y gracias a internet, la educación a distancia está cogiendo fuerza y ha empezado a demostrar su potencial. Tiene grandes ventajas como la flexibilidad proporcionada al estudiante, ya que este puede decidir el ritmo de aprendizaje según su disponibilidad, la gran cantidad de usuarios que se pueden inscribir al curso en un mismo año, etc... Como todo, también tiene desventajas y presenta algunos retos tecnológicos a los que hay que encontrar solución, seguridad y privacidad, escalabilidad debido a la gran cantidad de datos (Big Data [1]), problemas de seguimiento de los alumnos debido a la falta de contacto personal, etc...

Este TFG está centrado en la solución al problema de la gestión y el seguimiento de los alumnos para evitar posibles abandonos o la finalización insatisfactoriamente de los cursos. Para ello, se hará uso de las analíticas de aprendizaje (Learning Analytics) para recoger, analizar y extraer la información de las interacciones de los usuarios dentro de los cursos, con el objetivo de orientar a aquellos usuarios en riesgo de no finalizar un curso o de no obtener el certificado (los estudiantes que obtienen certificado es porque han finalizado el MOOC aprobándolo).

La Universidad Autónoma de Madrid (UAM) es miembro del consorcio edX [2] desde el año 2014, y ofrece MOOCs desde 2015. La gestión de estos cursos es llevada a cabo por la UAM, quien se encarga de anonimizar [3] los datos de los usuarios inscritos a dichos cursos. Esto ha permitido el desarrollo de una serie de herramientas para facilitar la visualización de datos y estadísticas y predecir que usuarios van a aprobar y cuáles no, gracias a la generación de indicadores y al aprendizaje automático.

Sobre el tema, se han presentado diferentes artículos en conferencias como el congreso LAK (Learning Analytics and Knowledge), la conferencia EDUCON (Global Engineering Education Conference), la cumbre eMOOCs (European MOOC Stakeholder Summit), la conferencia SIIE dentro del CEDI 2016 o la conferencia L@S (ACM Conference on Learning @ Scale).

Partimos de la herramienta *edX-MAS+*: *Model Analyzer System for edX MOOC*, lo que permitirá la generación de modelos predictivos en base a los indicadores importados en la plataforma.



## 1.2 Objetivos

El objetivo de este Trabajo de Fin de Grado es el desarrollo de una aplicación denominada edX-WS (edX Warning System). Esta herramienta permitirá el seguimiento, de manera sencilla, de los usuarios inscritos en MOOCs en edX mediante la visualización de una serie de indicadores y las estadísticas obtenidas de los datos anonimizados, la predicción de usuarios en riesgo de no terminar satisfactoriamente el curso y la generación de sugerencias para dichos usuarios en riesgo con el fin de facilitar a los creadores del curso (administradores y profesores) el encauzamiento de dichos usuarios hacia un final de curso exitoso.

Para la realización de este proyecto, se ha establecido los siguientes objetivos:

1. Entender la contextualización de los cursos MOOC y analíticas de aprendizaje a partir de la lectura de artículos.
2. Investigación y aprendizaje sobre librerías, lenguajes de programación y técnicas empleadas en este ámbito.
3. Análisis de los indicadores anonimizados facilitados por la UAM (<https://uam.es/UAM/MOOCs-en-edX/1446759793035.htm>).
4. Instalación y comprensión de la herramienta edX-MAS+.
5. Investigar la existencia de herramientas similares.
6. Definir requisitos funcionales y no funcionales que debe cumplir la aplicación.
7. Diseño de la aplicación mediante diagramas de clases.
8. Objetivos del desarrollo:
  - a. Uso de la herramienta edX-MAS+ para la generación de modelos predictivos.
  - b. Predicción y selección de usuarios en riesgo.
  - c. Visualización de todos los indicadores de cada uno de los usuarios en riesgo.
  - d. Exportación a CSV de los datos extraídos.
  - e. Visualización de comparativas de usuarios en riesgo con usuarios que terminaron satisfactoriamente el curso.
  - f. Generación de sugerencias basadas en los indicadores.
9. Integración de la funcionalidad descrita anteriormente en una aplicación web.
10. Pruebas de integración.
11. Estudio de posibles mejoras y trabajos a futuro que puedan resultar de este proyecto.
12. Redactar memoria de este Trabajo de Fin de Grado.

### 1.3 Organización de la memoria

La memoria consta de los siguientes capítulos:

- **Introducción:** Primer capítulo, en el que se realiza una breve introducción al proyecto. Las motivaciones que han dado lugar al desarrollo del mismo y los objetivos establecidos para su finalización.
- **Estado del arte:** Segundo capítulo, en el que se resume el estado actual de la educación a distancia y se expone el análisis realizado sobre la importancia y el potencial de los cursos MOOC, las predicciones basadas en modelos predictivos, la herramienta edX-MAS+.
- **Diseño:** Tercer capítulo, en el que se expone el análisis realizado para la elaboración del diseño de la aplicación (Arquitectura lógica, análisis de requisitos funcionales y no funcionales).
- **Desarrollo:** Cuarto capítulo, en el que se detallan las partes más importantes del desarrollo de esta aplicación.
- **Integración, pruebas y resultados:** Quinto capítulo, en el que se explican las pruebas realizadas para validar el resultado del desarrollo.
- **Conclusiones y trabajo futuro:** Sexto capítulo, el cual expone las conclusiones obtenidas del desarrollo de esta herramienta y los posibles trabajos a futuro para su mejora o ampliación.
- **Anexos:**
  1. **Anexo A:** Manual de Usuario
  2. **Anexo B:** Diagrama de Gantt



## 2 Estado del arte

---

En este capítulo se estudia la educación a distancia (*e-learning*), la importancia y el potencial los MOOCs, predicción en cursos online y un análisis de la herramienta edX-MAS+ y edX-WS.

### 2.1 Educación a distancia

La educación a distancia (o más conocido por su término en inglés, *e-learning*), ha permitido la difusión del conocimiento a nivel mundial gracias a las TIC [25]. Debido a sus múltiples ventajas como la eliminación de barreras geográficas, flexibilidad, costes, comodidad, etc. Se está convirtiendo en una opción de estudio cada vez más frecuente.

A continuación, se muestran datos recogidos por la organización Digital Learning Compass en su informe Distance Education Enrollment Report de 2017 [5]:

- Cada año 220.000 estudiantes apuestan por estudios en línea.
- Un 29.7% (aproximadamente 6.000.000) de los estudiantes, está inscrito en al menos un curso online.
- Las instituciones públicas abarcan el 67.8% de todos los estudiantes a distancia.
- Entre 2012 y 2015 el número de estudiantes que estudian en un campus ha disminuido casi un millón (931.317).

Los expertos en tecnología y educación ven la web como un gran sistema de información que permite hacer llegar recursos de aprendizaje a nivel mundial, con el propósito de crear experiencias de aprendizaje más enriquecedoras. El uso generalizado de internet en la enseñanza y en el aprendizaje ha dado lugar a la aparición de nuevas pedagogías y paradigmas de aprendizaje en los últimos años. En 2005, por ejemplo, Siemens propuso un nuevo modelo de aprendizaje para la era digital [20].

En este contexto aparecen los MOOC's (*Massive Online Open Courses*), de los cuales hablaremos a continuación.

### 2.2 Importancia y potencial de los MOOC's

Los MOOC [6] (*Massive Open Online Course*) son cursos masivos (debido a que están dirigidos a un número ilimitado de participantes) online.

Para que un curso sea considerado MOOC ha de cumplir ciertas características [7]:

- **Ser un curso.** Estructura orientada al aprendizaje y pruebas para acreditar el conocimiento adquirido por el estudiante.
- **Ser online.** El curso ha de poder ser impartido a distancia y usando internet como canal de comunicación. Permitiendo el acceso desde cualquier lugar con conexión a internet, de esta forma, no requiere presencia física en un aula.

- **Tener carácter masivo.** El principio, el límite de alumnos matriculados es ilimitado, o bien una cantidad muy superior a la de un curso presencial.
- **Alcance global.** Debe posibilitarse su acceso a través de internet independientemente de su ubicación geográfica.
- **Abierto.** Material y documentación han de ser accesibles de forma gratuita en internet.

El primer MOOC, ofrecido como tal de manera oficial, fue impartido por Stephen Downes y George Siemens en la universidad de Manitoba en Canadá en 2008 [6]. Abrieron al público un curso en línea titulado “Connectivism and Connective Knowledge CCK08”, al que se inscribieron 2.300 estudiantes.

En ese momento, Dave Cormier y Bryan Alexander propusieron “Massive Open Online Course” o MOOC.

Los MOOC dan respuesta a muchas de las necesidades demandas actualmente por la sociedad con respecto a la educación y a la formación.

A pesar de ofrecer grandes ventajas, también se presentan diversos problemas.

Tecnológicamente, los MOOC muestran problemas de escalabilidad debido a las grandes cantidades de alumnos inscritos en los cursos.

Educacionalmente, la tasa de abandono es muy alta, oscilando entre el 75% y el 95% [8]. Un estudio realizado en 2012 por C. O. Rodriguez [21], mostró que la tasa de abandono en los cursos de Stanford-AI era del 85%. Debido a que la mayoría de alumnos que se inscriben a cursos online, nunca participan o abandonan después de la primera visita [20]. Estos datos también encajan con los mostrados en una evaluación realizada a seis MOOC de la Universidad de Edimburgo, donde inicialmente 309.628 personas se inscribieron. Esta evaluación reveló que tan sólo un 29%, de los usuarios iniciales, llegaron a realizar el último examen (aunque curiosamente hubo una gran variación entre los seis cursos, oscilando estas cifras entre el 7% y el 59%) [24].

## 2.3 Predicción en cursos online

Debido al problema de la tasa de abandono y a la cantidad de alumnos que no obtienen el certificado en cursos online [21, 24], se han llevado cabo estudios para la detección de estos alumnos mediante modelos predictivos [26].

Para ello, se ha realizado un proceso de *Learning Analytics*, el cual consiste en [22]:

- **Recolección y pre-procesado de los datos:** La recolección de datos educativos es la base de los procesos de LA (*Learning Analytics*). La información recopilada puede ser demasiado grande o involucrar demasiados atributos irrelevantes, por lo que en ocasiones es necesario realizar un pre-procesamiento de los datos para transformarlos en un formato que sirva de entrada a un proceso de LA.
- **Análisis y acción:** A partir de los datos pre-procesados, se realizan diferentes acciones para obtener información valiosa y útil. Estas acciones incluyen monitoreo, análisis, predicción, intervención, evaluación, adaptación, personalización, recomendación y reflexión.
- **Post-procesamiento:** Puede implicar la adición de nuevas fuentes de datos, determinar nuevos atributos requeridos, identificar nuevos indicadores o métricas, modificar las variables o elegir un nuevo método de análisis.

Los conjuntos de datos se obtienen de diferentes indicadores como, por ejemplo:

- **Interacción en las plataformas en línea**
- **Indicadores de actividad**
- **Indicadores de material generado por los alumnos**
- **Indicadores particulares a los estudiantes**

Gracias a la elaboración de la herramienta edX-MAS+ podemos entrenar diferentes modelos para predecir el riesgo de los estudiantes de no obtener el certificado al finalizar el curso o de abandonarlo.

Los modelos escogidos para realizar estas predicciones son [9, 10]:

- Boosted Logistic Regression
- Stochastic Gradient Boosting
- k-Nearest Neighbors
- eXtreme Gradient Boosting
- Support Vector Machine
- Random Forest
- Naive Bayes
- Bayesian GLM
- CART Decision Tree
- Neuronal Network"

## **2.4 Test con modelos**

Para poder realizar predicciones con un modelo y poder obtener un valor estimado para una incógnita, será necesario entrenar (training) dicho modelo con una muestra de datos, de la cual conozcamos el valor real de la incógnita.

Una vez tenemos un modelo entrenado, podremos utilizarlo para realizar predicciones (dar un valor aproximado a la incógnita basándose en los datos con los que fue entrenado) es decir hacer test (testing) con una nueva muestra de datos.

La muestra de datos utilizada para entrenar el modelo ha de seguir la misma estructura que la muestra utilizada para el test.

La herramienta edX-WS utiliza modelos entrenados con cursos ya finalizados, para realizar el test con los indicadores de cursos que están transcurriendo. De esta forma podemos predecir el riesgo de no finalizar el curso o de no obtener el certificado para cada estudiante. El riesgo obtenido varía entre 0 y 1 (siendo 1 el máximo riesgo). La aplicación permite configurar los parámetros de entrada para el test y la elección del modelo (ya entrenado) con el que se realizará (Ver Anexo A).

Se emplea a usuarios que obtuvieron certificado (finalizando y aprobando el curso), como estudiantes de referencia para realizar comparaciones con los usuarios en riesgo (Véase Figura A-11).

## 2.5 Análisis de la herramienta edX-MAS+ y edX-WS

La herramienta edX-MAS+: Model Analyzer System para edX MOOC [9, 10], es una aplicación web que permite la generación de modelos predictivos diarios para un MOOC a partir de los indicadores obtenidos de los estudiantes. Gracias a estos modelos edX-WS puede realizar predicciones de usuarios en riesgos de no obtener el certificado y riesgo de abandonar el curso.

### 2.5.1 Leguajes de desarrollo y librerías

La herramienta edX-WS ha sido desarrollada con los lenguajes de programación R [11]. También hace uso de una base de datos relacional, PostgreSQL [19].



**Figura 2.5-1:**  
**Logo R**

R [11] es un lenguaje interpretado y *Open Source* enfocado al análisis estadístico. Es frecuente su uso en minería de datos, biomedicina, bioinformática y matemáticas financieras.

Gracias al desarrollo y a las aportaciones por parte de la comunidad, R superó en 2009 la cifra de 2000 paquetes (*CRAN*) en su repositorio oficial.

En este proyecto ha sido empleado para análisis y generación de modelos predictivos y para predicciones haciendo uso de los mismos. Para aportar más funcionalidad a la aplicación se ha hecho uso de varios paquetes:

- caret (Classification and Regression Training) [13]. Para generación de modelos predictivos y realización de predicciones.
- DT (DataTables) [14]. Wrapper sobre la librería *DataTables* de JavaScript, que añade más funcionalidad a las tablas generadas.
- ggplot2 [15]. Librería para la generación y visualización de gráficas.
- shiny [16]. Creación de aplicaciones web basadas en programación reactiva.
- shinydashboard [17]. Librería para la creación de Dashboards en R. Basada en Shiny.
- openxlsx [23]. Librería open source para escritura y lectura de documentos xlsx.



**Figura 2.5-2:**  
**Logo PostgreSQL**

PostgreSQL [19] es un sistema de gestión de bases de datos relacional, orientado a objetos y libre. Entre sus características, encontramos: alta concurrencia, amplia variedad de tipos nativos, disparadores (triggers), afirmaciones (assertions), vistas, integridad transaccional, herencia de tablas, tipos de datos y operaciones geométricas y transacciones distribuidas.

En este proyecto se ha hecho uso de esta base de datos para almacenar todos los indicadores de todos los cursos, los modelos predictivos generados, estadísticas de los modelos y las sugerencias.

### 2.5.2 Arquitectura de edX-MAS+

La aplicación edX-MAS+ consta de tres capas: **datos**, **negocio** y **presentación** [9, 10].

En la **capa de datos** tenemos PostgreSQL, que se encarga de almacenar toda la información y los datos de los cursos, así como los modelos generados, sus estadísticas y las sugerencias.

La **capa de negocio** es la encargada de procesar las peticiones recibidas por el usuario desde la capa de presentación y devolver una respuesta haciendo uso de la capa de datos durante el proceso. Esta capa incluye la generación de modelos, uso de los modelos en las predicciones, generación de gráficas y generación de ficheros para exportación.

La **capa de presentación** es la encargada de mostrar al usuario los datos procesados por la capa de negocio. Permite al usuario visualizar las gráficas generadas, exportar los ficheros generados, etc.

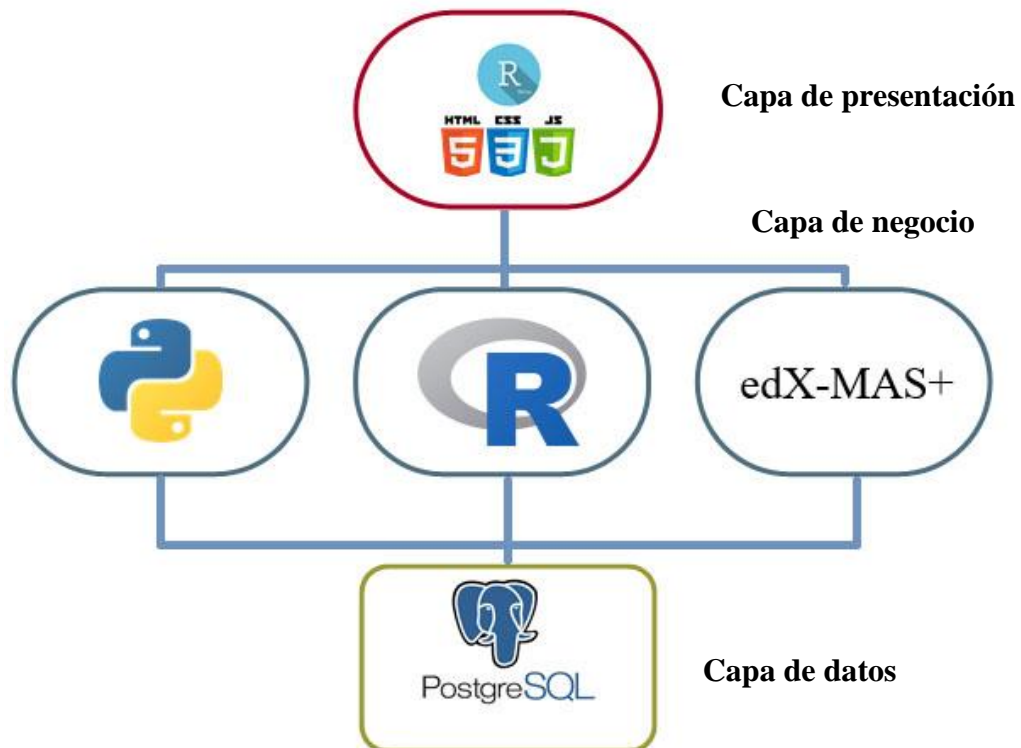


Figura 2.5-3: Arquitectura edX-MAS+



### 2.5.3 Indicadores disponibles

Gracias a la integración realizada con edX-MAS+ es posible realizar predicciones con los siguientes indicadores, obtenidos cada día para cada usuario [9, 10, 27]:

- Número total de eventos generados por el alumno (**num\_events**).
  - Número total de sesiones del alumno durante el día (**num\_sessions**).
  - Tiempo total invertido en un curso (**total\_time**).
  - Número total (**nav\_events**) y tiempo (**nav\_time**) entre eventos relacionados con la navegación a través del curso.
  - Número y tiempo entre eventos relacionados con las interacciones de vídeo (**video\_events**, **video\_time**).
  - Número y tiempo entre eventos del foro del curso (**forum\_events**, **forum\_time**).
  - Número total de interacciones con los problemas del curso y tiempo entre dichas interacciones (**problem\_events**, **problem\_time**).
- 
- Número de días conectado (**connected\_days**).
  - Número de días de inactividad consecutivos (**consecutive\_inactivity\_days**).
  - Número de vídeos distintos accedidos (**num\_diff\_videos**).
  - Número de problemas distintos accedidos (**num\_diff\_problems**).



## 3 Diseño

Se ha realizado un diseño basado en lo establecido previamente por las herramientas edX-MAS+, así como por lo definido en los requisitos funcionales que ha de tener la aplicación edX-WS. También detallaremos los diferentes módulos que conforman la herramienta.

### 3.1 Arquitectura Lógica

La arquitectura lógica consiste en un conjunto de patrones que proporcionan un diseño de alto nivel en que se pueden observar los componentes principales y sus interrelaciones.

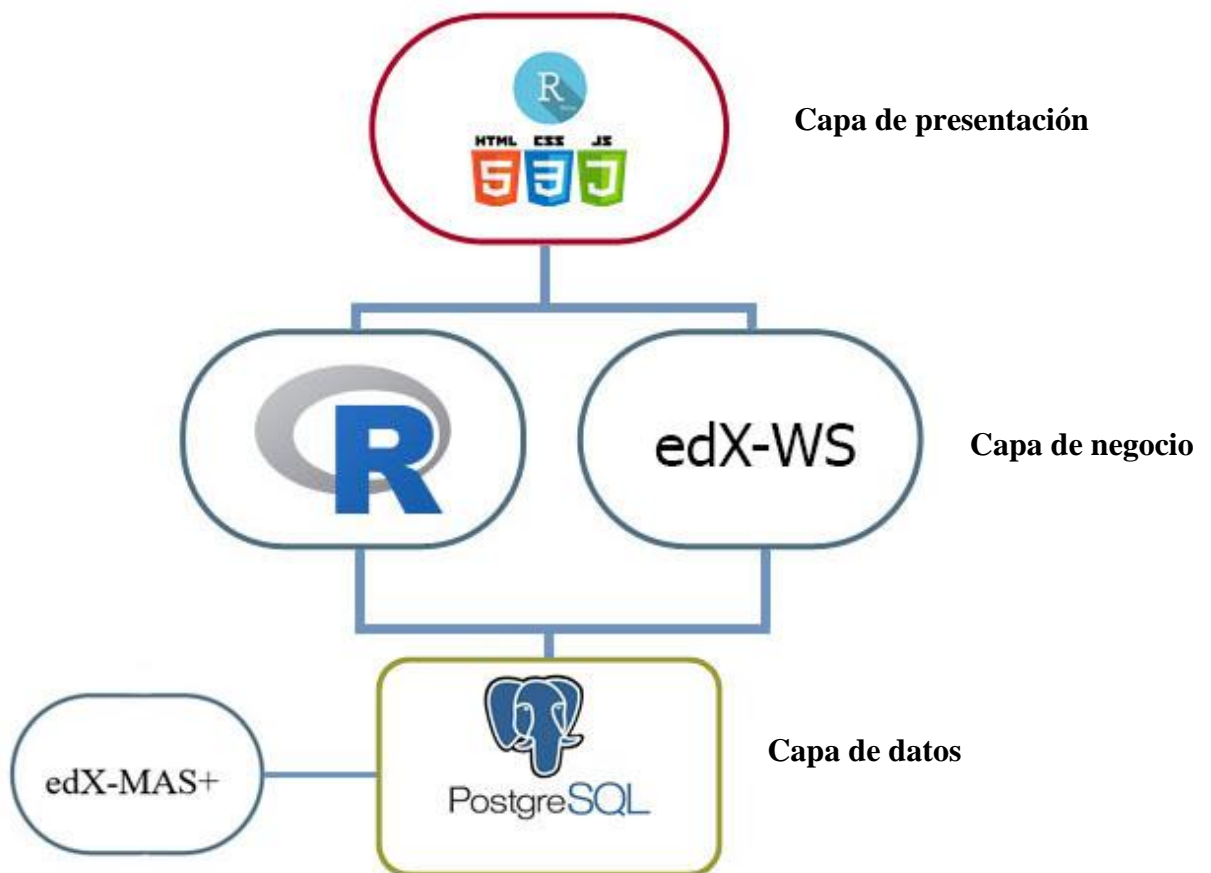


Figura 3.1-1: Arquitectura lógica de edX-WS

### **3.1.1 Capa de presentación**

También denominada “interfaz de usuario”, es la capa encargada de mostrar la información al usuario, de recoger la información introducida por el mismo para su procesamiento en la capa de negocio. Ha de ser “user-friendly” (fácil de usar).

Debido a que en este caso edX-WS es una aplicación web, esta capa emplea HTML5 y CSS gracias al framework Shiny de R [16].

### **3.1.2 Capa de negocio**

Es la capa encargada de procesar las peticiones realizadas por el usuario desde la capa de presentación y dar una respuesta del resultado del proceso. También se comunica con la capa de datos para almacenar o recuperar información de la base de datos.

La capa de negocio de edX-WS está desarrollada en el lenguaje de programación R y se integra con la herramienta edX-MAS+. Esta capa contiene la funcionalidad más importante, como la utilización de modelos predictivos para realizar test sobre los datos de los usuarios, generación de sugerencias, generación de ficheros de exportación, generación de gráficas para la capa de presentación...

### **3.1.3 Capa de datos**

Es la capa donde se almacenan los datos generados por la aplicación. La herramienta edX-WS utiliza el gestor de base de datos PostgreSQL para almacenar y recuperar la información generada por la capa de negocio.

### 3.2 Requisitos funcionales

RF-01: Generación de modelos predictivos	
Actor	Usuario
Descripción	El sistema debe permitir generar modelos predictivos
Precondiciones	El usuario ha cargado todos los indicadores de todos los días del curso que se va a emplear para la generación del modelo (training). El usuario ha abierto la aplicación.
Post-condiciones	El sistema almacena el modelo generado
Escenario Principal	
1.- El usuario entra en la pantalla de creación de modelos (“Create Models”)	
2.- El usuario selecciona el curso, la edición, los indicadores, el algoritmo con el que hacer entrenamiento (training) y crear el modelo predictivo, la salida a predecir (abandono o dropout / obtención de certificado) y la frecuencia (modelo diario o modelo semanal).	
3.- El usuario pulsa el botón “Get Model”. Esto lanza un proceso que genera un modelo de predicción de la salida indicada (output) basado en el algoritmo seleccionado para la frecuencia indicada con los indicadores de entrada (inputs) establecidos previamente. Cuando el proceso finaliza, almacena el modelo generado en la base de datos (En este caso PostgreSQL).	
4. Cuando finaliza el proceso el sistema muestra en la tabla de modelos el modelo generado.	

Véase Figura A-1, Figura A-2, Figura A-3 y Figura A-4 del anexo A

Tabla 3-1: RF-01

<b>RF-02: Test con modelos predictivos</b>	
<b>Actor</b>	Usuario
<b>Descripción</b>	El sistema debe permitir realizar <b>test</b> con los modelos predictivos cargados en el sistema.
<b>Precondiciones</b>	El usuario ha abierto la aplicación. El usuario ha generado un modelo predictivo. El usuario ha cargado en el sistema los datos del nuevo curso contra el que se hará el test.
<b>Post-condiciones</b>	El sistema carga las tablas con los datos resultantes del proceso de test
<b>Escenario Principal</b>	
<b>1.- El usuario entra en la pantalla de creación de modelos (“Test Models”).</b>	
<b>2.- El usuario selecciona un modelo. Para ello selecciona el curso, la edición, algoritmo, indicadores y salida del curso (dropout / certificado) utilizado para generar el modelo predictivo, es decir, con el que se hizo training.</b>	
<b>3.- El usuario selecciona el curso, la edición, algoritmo, los indicadores y día límite para hacer el test, es decir, se selecciona el nuevo conjunto de datos con el que hacer testing.</b>	
<b>4.- El usuario pulsa el botón “Test”. Esto lanza un proceso que realiza un test con los nuevos datos de entrada y el modelo seleccionado (previamente entrenado con datos de cursos anteriores). El cual, tras finalizar muestra en una tabla los resultados de los usuarios en riesgo. A su vez, genera todos los datos necesarios para exportaciones y sugerencias.</b>	

Véase Figura A-5, Figura A-6, Figura A-7 y Figura A-8 del anexo A

**Tabla 3-2: RF-02**

<b>RF-03: Generar sugerencias para usuarios en riesgo</b>	
<b>Actor</b>	Usuario
<b>Descripción</b>	El sistema debe permitir generar sugerencias para usuarios en riesgo
<b>Precondiciones</b>	El usuario ha abierto la aplicación. El usuario ha realizado un test con un modelo predictivo.
<b>Post-condiciones</b>	El sistema genera la información necesaria para poder exportar a CSV las sugerencias de los usuarios en riesgo.
<b>Escenario Principal</b>	
<b>1.- El usuario entra en la pantalla de creación de modelos (“Test Models”).</b>	
<b>2.- El usuario pulsa el botón “Download csv suggestions”. Con los datos obtenidos después de realizar el test, se obtienen las sugerencias de la base de datos y se le asignan a los usuarios en base a sus indicadores.</b>	

Véase Figura A-10 del anexo A

**Tabla 3-3: RF-03**

<b>RF-04: Exportación de los datos obtenidos del test</b>	
<b>Actor</b>	Usuario
<b>Descripción</b>	El sistema debe permitir exportar los datos obtenidos del test.
<b>Precondiciones</b>	El usuario ha abierto la aplicación. El usuario ha realizado un test con un modelo predictivo.
<b>Post-condiciones</b>	El sistema genera la información necesaria para poder exportar a CSV los datos resultantes del test.
<b>Escenario Principal</b>	
<b>1.- El usuario entra en la pantalla de creación de modelos (“Test Models”).</b>	
<b>2.- El usuario pulsa uno de los botones “Download summary of users without risk”, “Download csv of users with risk” o “Download csv of all users” para obtener los diferentes ficheros de exportación.</b>	

Véase Figura A-10 del anexo A

**Tabla 3-4: RF-04**

<b>RF-05: Generación de gráficas comparativas para cada usuario</b>	
<b>Actor</b>	Usuario
<b>Descripción</b>	El sistema debe permitir generar gráficas para comparar los indicadores del usuario con el valor medio de los mismos de usuarios que obtuvieron el certificado.
<b>Precondiciones</b>	El usuario ha abierto la aplicación. El usuario ha realizado un test con un modelo predictivo.
<b>Post-condiciones</b>	El sistema genera y muestra las gráficas comparativas para el usuario seleccionado.
<b>Escenario Principal</b>	
<b>1.- El usuario entra en la pantalla de creación de modelos (“Test Models”).</b>	
<b>2.- El usuario selecciona en la tabla resultante de test, el estudiante del cual quiere ver sus comparativas.</b>	
<b>3.- El usuario entra en la pantalla de “User Details”. Gracias a los resultados obtenidos en el test y los indicadores de usuarios de cursos previos (en este caso del curso con el que el modelo fue entrenado), se dibujan gráficas comparativas.</b>	

Véase Figura A-11 del anexo A

**Tabla 3-5: RF-05**

<b>RF-06: Generación de gráficas comparativas para grupos de usuarios</b>	
<b>Actor</b>	Usuario
<b>Descripción</b>	El sistema debe permitir generar gráficas para comparar los indicadores de los usuarios en riesgo con los indicadores de usuarios que obtuvieron el certificado.
<b>Precondiciones</b>	El usuario ha abierto la aplicación. El usuario ha realizado un test con un modelo predictivo.
<b>Post-condiciones</b>	El sistema genera y muestra las gráficas comparativas para los grupos de usuarios.
<b>Escenario Principal</b>	
<b>1.- El usuario entra en la pantalla de “Group Details”</b>	
<b>2.- El usuario selecciona en el desplegable el indicador que desea comparar. Gracias a los resultados obtenidos en el test y los indicadores de usuarios de cursos previos (en este caso del curso con el que el modelo fue entrenado), se dibujan gráficas comparativas (boxplots).</b>	

Véase Figura A-12 del anexo A

**Tabla 3-6: RF-06**



### 3.3 Requisitos no funcionales

RNF-01: Usabilidad	
Descripción	Debe ser fácil de usar. Con ayudas e interfaces intuitivas.
Prioridad	5
RNF-02: Seguridad	
Descripción	El sistema debe mantener de manera anónima la identidad de los usuarios, identificándolos con un id. Esto se consigue anonimizando los datos.
Prioridad	5
RNF-03: Multiplataforma	
Descripción	La aplicación deberá funcionar en distintos tipos de sistemas operativos y plataformas de hardware.
Prioridad	3
RNF-04: Rendimiento	
Descripción	El sistema debe soportar el manejo de gran cantidad de información durante su proceso.
Prioridad	3
RNF-05: Escalabilidad	
Descripción	El sistema debe permitir añadir fácilmente nuevos indicadores, modelos predictivos, cambiar y añadir nuevas sugerencias, añadir nuevas métricas...
Prioridad	4

Tabla 3-7: Requisitos no funcionales

### 3.4 Diseño de los módulos

Una vez establecidos los requisitos funcionales y haber realizado un estudio sobre el estado del arte, se han definido los diferentes módulos que conforman la aplicación edX-WS. De esta manera, cada módulo contiene una funcionalidad concreta y acotada, lo que permite escalar esta aplicación cuando requiera ser ampliada o cuando se use en otros proyectos y que el mantenimiento sea menos costoso.

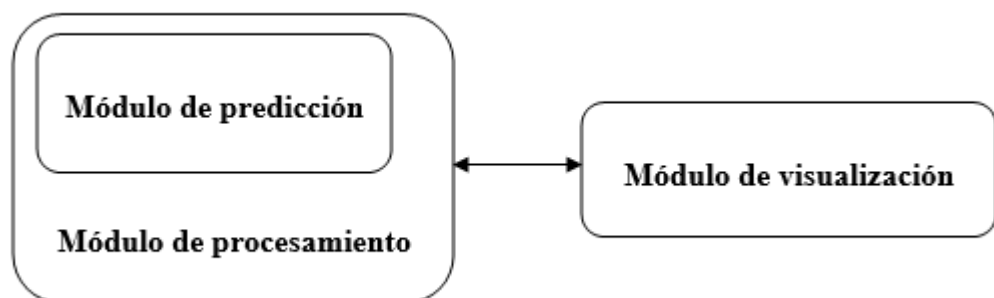


Figura 3.4-1: Módulos edX-WS

### 3.4.1 Módulo de predicción

Este módulo es el encargado de hacer uso de los modelos predictivos generados con la librería *caret* [13]. Estos modelos están entrenados a partir del conjunto de indicadores de todos los usuarios de un curso ya finalizado. Esta parte es la denominada como *training*.

Haciendo uso de la función *predict* de la librería estándar de R [11], podemos realizar predicciones sobre conjuntos de indicadores pertenecientes a estudiantes que están inscritos en cursos MOOC. Esta parte es la denominada como *testing*, como ya se ha mencionado en apartados anteriores.

Estas predicciones nos dicen el riesgo que tiene un estudiante de abandonar el curso o de no obtener certificado al finalizarlo (dependiendo de los parámetros establecidos al realizar las predicciones). El valor del riesgo varía entre 0 y 1 (siendo 0, la ausencia total de riesgo y 1, el riesgo absoluto).

### 3.4.2 Módulo de procesamiento

Este módulo se encarga de realizar los procesos necesarios para extraer todos los indicadores de los usuarios en riesgo y de los usuarios que no están en riesgo, a partir de la información obtenida del módulo de predicción.

Todos estos datos, servirán para nutrir al módulo de visualización y exportación, permitiendo la generación de ficheros *csv* y *xlsx*, y la generación de gráficas comparativas, tanto a nivel de usuario, comparando los indicadores de un usuario en riesgo con la media de los valores de cada indicador de los usuarios que obtuvieron certificado, como a nivel de grupo, pudiendo comparar los valores de los indicadores de todos los usuarios en riesgo con los indicadores de los usuarios que finalizaron y obtuvieron el certificado. También es el encargado de la generación de sugerencias mediante la clasificación de cada indicador con un nivel de gravedad, de esta forma, se comparan los indicadores de cada usuario en riesgo con un *summary* de los usuarios que no están en riesgo y dependiendo del valor del indicador, la sugerencia tomará un valor de gravedad.

Cabe destacar que los procesos de este módulo tienen una gran carga computacional debido a la cantidad masiva de datos.

### 3.4.3 Módulo de visualización y exportación

Este módulo es el encargado de representar visualmente todos los componentes de la aplicación y permitir la comunicación entre la capa de presentación y la capa lógica del sistema.

Se encarga de representar una interfaz gráfica para:

- La generación de modelos predictivos, pudiendo seleccionar todos los parámetros necesarios para ello, así como representar en una tabla todos los modelos ya cargados en el sistema.
- Visualizar los datos obtenidos al realizar el *testing* con el modelo predictivo y la configuración necesaria para realizarlo.
- Descarga de ficheros de exportación *csv* generados por el módulo de procesamiento.
- Visualización de las gráficas comparativas de cada usuario generadas por el módulo de procesamiento.

- Descarga del fichero de exportación *xlsx* con los datos de las gráficas de usuario.
- Visualización de las gráficas a nivel grupal generadas por el módulo de procesamiento.
- Descarga del fichero de exportación *xlsx* con los datos de las gráficas de los grupos de usuarios.



## 4 Desarrollo

El desarrollo de la aplicación se ha realizado siguiendo la arquitectura lógica y los requisitos (funcionales y no funcionales) descritos previamente. En este apartado se detalla la estructura de ficheros y los puntos más importantes de cada uno de ellos.

### 4.1 Estructura de los ficheros

/postgresqlstore/	/R/	/R/ws/
createdatabase.sql	edxpostgresqlb.R	server.R
edxindicatorstore.py	predictor.R	ui.R
edxpostgresqlstore.py		
script_all_course.py		

Tabla 4-1: Estructura de ficheros

#### 4.1.1 Directorio /postgresqlstore

En este directorio se encuentran los ficheros para importar los datos de los nuevos cursos. Almacenando su información y todos sus indicadores.

- createdatabase.sql: Contiene en formato sql la estructura de la base de datos y la definición de las tablas.
- edxindicatorstore.py: Contiene las clases y los métodos para la generación de indicadores.
- edxpostgresqlstore.py: Contiene la lógica necesaria para realizar la importación de cursos. Inserta en la base de datos toda la información del curso.
- script\_all\_course.py: A partir de una ruta (donde están ubicados los datos del curso), el nombre del curso, el nombre de la edición y una lista de indicadores, guarda los datos del curso y todos sus indicadores.

#### 4.1.2 Directorio /R

En este directorio se encuentra el fichero que sirve de interfaz con la base de datos y el fichero necesario para realizar las predicciones con los modelos predictivos.

- edxpostgresqlb.R: Es el fichero encargado de servir de interfaz para la comunicación con la base de datos. Contiene el código necesario para realizar la conexión con la base de datos y las consultas de tipo *SELECT* necesarias para obtener la información de los cursos, los modelos predictivos, los indicadores...
- predictor.R: Contiene el código necesario para la generación de modelos predictivos y el uso de los mismos para realizar predicciones.

### 4.1.3 Directorio /R/ws

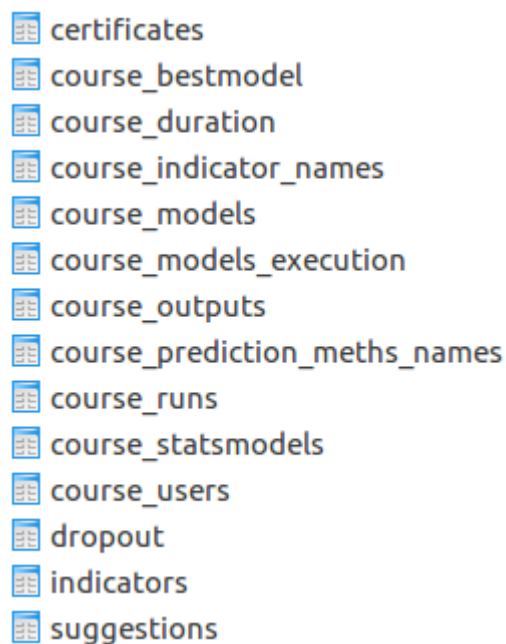
En este directorio se encuentran los ficheros de la capa de negocio y de la capa de presentación de la aplicación edX-WS, ambos desarrollados en el lenguaje de programación R.

- server.R: Es el fichero que contiene toda la lógica de la capa de negocio. Sirve de controlador para dar respuesta a la capa de presentación. Siguiendo la estructura de una aplicación Shiny, es la parte del servidor.
- ui.R: Es el fichero encargado de la capa de presentación. Permite al usuario enviar peticiones a la capa de negocio y se ocupa de mostrar los resultados. Siguiendo la estructura de una aplicación Shiny, es la parte de la interfaz de usuario. El código desarrollado en R con el framework Shiny, genera código HTML5 para crear una aplicación web que pueda ser accedida mediante navegador.

## 4.2 Estructura de la Base de Datos

La base de datos se compone de 14 tablas que siguen modelo relacional. La base de datos que utiliza edX-WS es PostgreSQL [19].

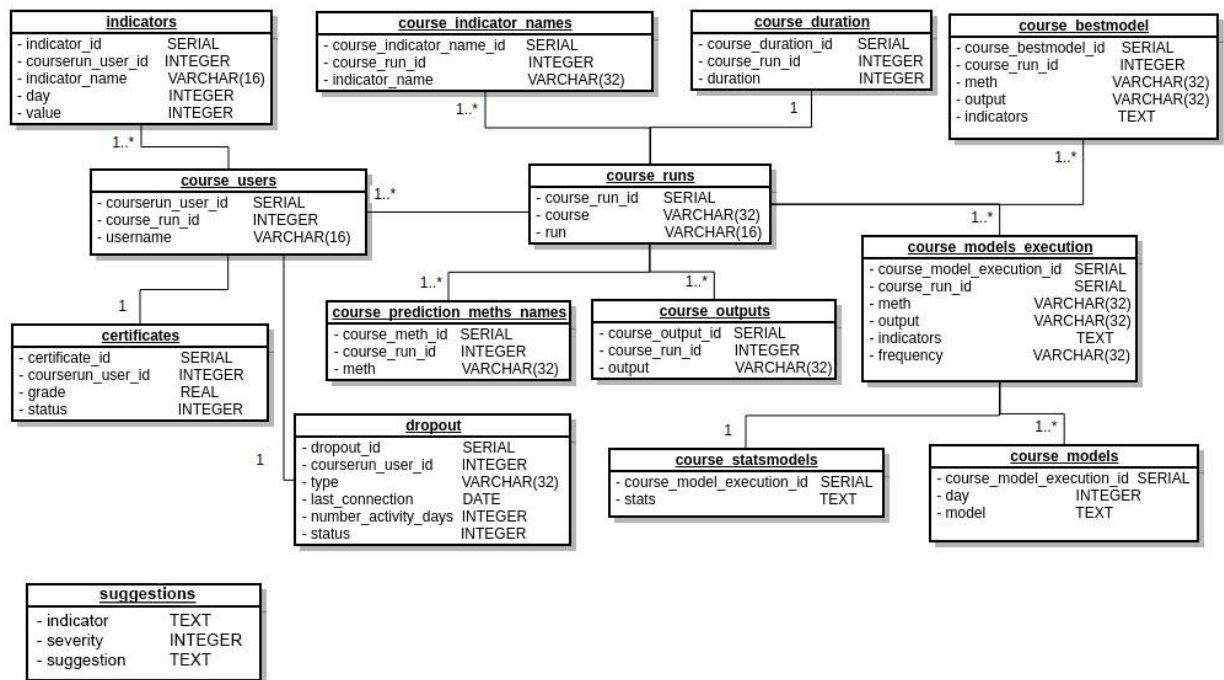
A continuación, se muestra un listado de las tablas, una breve explicación y las relaciones entre cada una de ellas.



**Figura 4.2-1: Tablas de la base de datos**

- Certificates: Tabla donde se almacenan los certificados para cada edición.
- Course\_bestmodel: Tabla donde se almacena el mejor modelo para cada edición de cada curso.
- Course\_duration: Tabla donde se almacena la duración (en días) de cada curso.
- Course\_indicator\_names: Tabla donde se almacenan los nombres de todos los indicadores disponibles para una edición de un curso.

- **Course\_models:** Tabla donde se almacenan los modelos ya generados y entrenados.
- **Course\_models\_execution:** Tabla donde se almacena la información relativa a los modelos generados (curso, edición, algoritmo, output, indicadores y frecuencia).
- **Course\_outputs:** Tabla donde se almacenan los outputs disponibles para cada edición.
- **Course\_prediction\_meth\_names:** Tabla que almacena los nombres de todos los algoritmos disponibles para cada edición de un curso.
- **Course\_runs:** Tabla donde se almacenan las ediciones disponibles para cada curso.
- **Course\_statsmodels:** Tabla donde se almacenan estadísticas para los modelos predictivos generados.
- **Course\_users:** Tabla donde se almacenan los usuarios inscritos para cada edición.
- **Dropout:** Tabla donde se almacena la información de abandono para cada usuario.
- **Suggestions:** Tabla donde se almacenan los mensajes para la generación de sugerencias.



**Figura 4.2-2: Modelo de la base de datos**

## 4.3 Implementación

A continuación, se explica cómo se han implementado las funcionalidades más importantes de la aplicación edX-WS.

### 4.3.1 ui.R

Interfaz de usuario (capa de presentación) desarrollada con la librería Shiny Dashboard (basada en el framework Shiny). Se compone de un menú lateral (*sidebarMenu*) y la página de contenido (*dashboardBody*).

- Menú lateral: Compuesto por diferentes componentes como despleables (*selectInput*), barras de selección (*sliders*), radio buttons, check boxes y botones.
- Página de contenido: Parte de la interfaz gráfica donde se renderizan los resultados de la capa de negocio. Representando los resultados con gráficas (*ggplot2*) y con tablas (*DT*). Se ha empleado la librería *DT* para representar tablas, ya que las tablas por defecto de Shiny no son seleccionables, por el contrario, las tablas construidas con la librería *DT* permiten seleccionar las filas. Funcionalidad necesaria para la pantalla de detalle de usuario (*User Details*, ver Anexo A).

### 4.3.2 server.R

Servidor basado en el framework Shiny. Es el encargado de recibir las peticiones de la capa de presentación y dar respuesta a las mismas. Utiliza las librerías *ggplot2* para la generación de gráficas, *DT* para la generación de tablas, *shinyDashboard* para la generación de componentes para la capa de presentación, *caret* para realizar las predicciones con los modelos predictivos. Hace uso de los ficheros *edxpostgreslodb.R* y *predictor.R* mencionados anteriormente.

A continuación, se describen las funciones implementadas en este fichero:

- Las funciones *modelsTable*, *courseModel\_selectInput*, *runModel\_radioButtonsInput*, *outputModel\_selectInput*, *methModel\_selectInput*, *getFrequency\_selectInput*, *getIndicators\_checkBoxesInput*, *getModel\_actionButtonInput* son necesarias para definir los componentes necesarios para el menú lateral de la pantalla *Create Models* (ver Anexo A).
- El **callback** definido para el botón *getModelButton*, utiliza funciones definidas en el fichero *predictor.R* para la generación del modelo predictivo basado en los parámetros de configuración establecidos en el menú lateral de la pantalla *Create Models*.
- Las funciones *modelCoursesTest\_selectInput*, *modelRunsTest\_radioButtonsInput*, *modelMethsTest\_selectInput*, *modelOutputTest\_selectInput*, *courseTest\_selectInput*, *runTest\_radioButtonsInput*, *getIndicatorsTest\_checkBoxesInput*, *getIndicatorsTest\_selectInput*, *testButton\_actionButtonInput* son necesarias para definir los componentes necesarios para el menú lateral de la pantalla *Test Models* (ver Anexo A).



- El **callback** definido para el botón **testButton** ejecuta la predicción con el modelo seleccionado y los datos del nuevo curso seleccionados en el menú lateral.  
Esta función también establece cuales son los usuarios con riesgo de no finalizar el curso basándose en los resultados de la predicción. Para ello, realiza un test previo (función *getModelMeanRisk*) con los datos con los que se generó el modelo, obteniendo la media de riesgo, pero conociendo los resultados reales de los estudiantes. Posteriormente se realiza el test con los nuevos datos del nuevo curso y marca a aquellos estudiantes que estén por encima de la media de riesgo como posibles usuarios con peligro de no finalizar el curso. Como resultado de esta petición el servidor devuelve a la capa de presentación una tabla con los resultados, para ello se ha hecho uso de la librería *DT*.
- Las funciones *callback downloadNonRiskSummaryCsv*, *downloadCsv*, *downloadAllCsv*, *downloadSuggestionsCsv* son las encargadas de permitir la exportación de los datos generados en el test.
- La función **getSuggestions**, es la encargada de la generación del fichero de sugerencias. Para ello evalúa cada uno de los indicadores de cada usuario y asignándole una gravedad, comparando el valor del indicador con la mediana y el primer cuartil de los resultados de los estudiantes que obtuvieron certificado. Este índice de gravedad es traducido a una sugerencia consultando en la base de datos la tabla *suggestions*. De este modo, pueden añadirse o modificarse de manera sencilla los índices de gravedad y los mensajes de sugerencias.
- Debido a que, en una aplicación web por cada petición, el servidor sólo puede dar una respuesta se han desarrollado las funciones **plots** y **calculatePlots**. Se encargan de generar varias gráficas para cada indicador, de este modo se devuelven varias gráficas en una única respuesta. Para la generación de estas gráficas se ha hecho uso de *ggplot2*. Estas gráficas se visualizan en la pantalla *User Details* (ver Anexo A).
- Las funciones *indicatorsWithoutRiskBoxplot* y *indicatorsWithRiskBoxplot* son las necesarias para la generación de las gráficas boxplot para la comparación de grupos. Estas gráficas son renderizadas en la pantalla de *Group Details* (ver Anexo A).



## 5 Integración, pruebas y resultados

---

Para validar un correcto funcionamiento del software y garantizar su calidad, se han realizado diferentes tipos de pruebas (tanto funcionales como no funcionales). Para ello, se han empleado datos reales anonimizados de cursos MOOCs de la UAM.

### 5.1 Pruebas funcionales

Se trata de pruebas basadas en ejecución, revisión y retroalimentación. Son pruebas específicas para validar que el software funciona correctamente.

#### 5.1.1 Pruebas unitarias

Son pruebas para validar el correcto funcionamiento de una unidad de código. De esta manera se prueba la funcionalidad independientemente del resto de la aplicación, validando cada fichero individualmente.

- Se han realizado pruebas sobre la generación de los modelos predictivos, verificando que los modelos diarios se almacenan correctamente en la base de datos. Se comprueba que el modelo se ha generado en base a los parámetros de configuración especificados.
- Se han realizado pruebas sobre la predicción usando los modelos previamente generados. Para ello se ha comprobado que los modelos se pueden obtener de la base de datos, que son válidos para la librería *caret*, que se pueden utilizar contra cualquier conjunto de datos nuevo que comparta la misma estructura que el conjunto de datos con el que se generó el modelo. También se ha verificado que los datos obtenidos en la predicción son coherentes con los datos utilizados para el testing.
- Se ha verificado la correcta generación de los ficheros, así como la posibilidad de ser descargados.
- En cuanto a la capa de presentación se ha probado que todos los resultados y componentes devueltos por el servidor se renderizan correctamente en la interfaz de usuario.

#### 5.1.2 Pruebas de integración

Pruebas que validan que todas las unidades de código funcionan correctamente de manera conjunta.

- Se han realizado pruebas para verificar que las diferentes funciones de la capa de negocio que necesitan acceder a la capa de datos pueden hacerlo de manera correcta.
- Se ha verificado que los datos generados por la herramienta edX-MAS+ legibles por la aplicación edX-WS. Permitiendo generar modelos predictivos.
- Se ha verificado que los modelos generados previamente son válidos y aptos para su uso en la predicción.
- Se han validado las gráficas obtenidas a partir de los datos generados en las predicciones.

### 5.1.3 Pruebas de aceptación

Son pruebas para determinar si el software cumple con los requisitos especificados por el usuario o empresa que lo solicita.

Para ello se ha instalado la aplicación en el sistema de un investigador de la institución, para verificar el correcto funcionamiento de la aplicación y valorar su utilidad.

El test realizado por el investigador fue el siguiente:

1. El usuario presiona el botón “*Test*” eligiendo un modelo ya generado (para el curso Quijote501x, edición 1T-2015 y con output certificado), para el curso Quijote501x, edición 3T-2016, con output de certificado y para 25 días.
2. El usuario presiona el botón “*Show Results*” y se muestran los resultados del test.
3. Exporta todos los datos, descargando todos los ficheros *csv* (Ver figura 5.1-1).
4. Examina y analiza los datos de los ficheros. Los datos encajan con lo esperado.
5. Selecciona el usuario con id 1843647 para realizar el análisis. Es un usuario en riesgo.
6. El usuario va a la pantalla “*User Details*”, exporta los datos de las gráficas a un fichero de Excel.
7. Abre sin problema los ficheros exportados con Excel.
8. Genera gráficas Excel con los datos exportados y encajan con las generadas por la aplicación. Se observa que el usuario comienza a dejar de tener actividad en torno al día 15
9. El usuario va a la pantalla “*Group Details*”, exporta los datos de las gráficas a un fichero de Excel.
10. Abre sin problema los ficheros exportados con Excel.
11. Genera gráficas Excel con los datos exportados y encajan con las generadas por la aplicación.
12. Extrae conclusiones de las gráficas: Los usuarios en riesgo hacen pocos problemas y no entran al foro, pero si ven algún video y navegan unos días por el curso. Las gráficas de días inactivos consecutivos se invierten (los usuarios en riesgo tienen valores más altos que los usuarios sin riesgo) según lo previsto.
13. Se realiza un nuevo test con los mismos parámetros, pero para 10 días y se busca de nuevo al usuario 1843647.
14. Se descarga de nuevo todos los ficheros de exportación.
15. Se observa que a esas alturas, el usuario no estaba en riesgo puesto que aún realiza tareas, estaba activo en el curso.
16. Se realiza un nuevo test con los mismos parámetros, pero para 5 días.
17. Se descarga todos los ficheros de exportación para el nuevo test.
18. Se valida que los datos siguen siendo coherentes.

Nombre	Tamaño
AllUsers-Quijote501x-5.csv	23 KB
AllUsers-Quijote501x-10.csv	31 KB
AllUsers-Quijote501x-25.csv	42 KB
GroupDetails-Quijote501x-consecutive_i...	24 KB
GroupDetails-Quijote501x-num_diff_pro...	18 KB
GroupDetails-Quijote501x-num_diff_vide...	14 KB
GroupDetails-Quijote501x-num_events-2...	19 KB
GroupDetails-Quijote501x-video_events-...	19 KB
GroupDetails-Quijote501x-video_time-25...	19 KB
UserDetails-1843647-Quijote501x-5.xlsx	18 KB
UserDetails-1843647-Quijote501x-10.xlsx	19 KB
UserDetails-1843647-Quijote501x-25.xlsx	29 KB
UserDetails-7532637-Quijote501x-25.xlsx	29 KB
NonRiskUsersSummary-Quijote501x-5.csv	1 KB
NonRiskUsersSummary-Quijote501x-10.c...	1 KB
NonRiskUsersSummary-Quijote501x-25.c...	1 KB
Suggestions-Quijote501x-5.csv	42 KB
Suggestions-Quijote501x-10(1).csv	152 KB
Suggestions-Quijote501x-25.csv	218 KB
UsersWithoutRisk-Quijote501x-5.csv	12 KB
UsersWithoutRisk-Quijote501x-10.csv	5 KB
UsersWithoutRisk-Quijote501x-25.csv	4 KB
UsersWithRisk-Quijote501x-5.csv	12 KB
UsersWithRisk-Quijote501x-10.csv	27 KB
UsersWithRisk-Quijote501x-25.csv	38 KB

**Figura 5.1-1: Ficheros exportados**

## 5.2 Pruebas no funcionales

Son pruebas cuyo objetivo es validar los requisitos no funcionales especificados en el diseño de la aplicación.

### 5.2.1 Pruebas de compatibilidad

Son pruebas que verifican la correcta funcionalidad del software en diferentes entornos.

La aplicación edX-WS ha sido probada en dos equipos con diferentes sistemas operativos (Linux y Windows), con prestaciones (diferente memoria RAM, diferente procesador...).



## 6 Conclusiones y trabajo futuro

---

A continuación, se exponen las conclusiones obtenidas de la realización de este proyecto y los posibles trabajos a futuro a realizar sobre este proyecto para mejorarlo.

### 6.1 Conclusiones

Se propone una aplicación, la cual se ha desarrollado basándose en técnicas de *Learning Analytics* para predecir el fracaso de los estudiantes en cursos MOOC. Para ello, el conjunto de herramientas cubre toda la secuencia de pasos, desde la generación de modelos predictivos, su uso para realizar predicciones, hasta la generación de ficheros de exportación de los usuarios en riesgo y sugerencias. Estas sugerencias, pretenden facilitar a los investigadores que trabajan con datos de los MOOCs la tarea de seguimiento y ayuda al estudiante.

Para la realización de la aplicación edX-WS, se han completado las metas y objetivos definidos en este documento y que a continuación se detallan.

En una primera versión se realizó un estudio de la herramienta edX-MAS+, analizando las tecnologías empleadas y la infraestructura utilizada. Todo esto se ha tenido en cuenta en el diseño de edX-WS para maximizar la compatibilidad y facilitar el uso de datos generados por edX-MAS+. Posteriormente se hizo un estudio de las librerías utilizadas para la generación de modelos y un aprendizaje del contexto de la plataforma. Debido a que es una aplicación basada en *Learning Analytics* y es necesario hacer uso de herramientas estadísticas, predicción, generación de gráficas, etc. Se optó por emplear el lenguaje de programación R, el cual ofrece técnicas de análisis de datos, generación de gráficos y cuenta con una comunidad de usuarios muy activa.

Tras este estudio se marcaron unas metas, que consistían en uso de los datos generados por edX-MAS+ por parte de edX-WS, la posibilidad de utilizar los modelos en predicciones, identificación de los usuarios en riesgo, exportación de los resultados, generación de sugerencias y obtención de gráficas comparativas. Para el desarrollo de todas estas funcionalidades, también se realizó un análisis de las librerías que más se ajustaban a los requisitos establecidos, siendo necesario hacer uso de bibliotecas para la exportación de ficheros en formato *xlsx*, para la generación de gráficas, etc.

Gran parte de los esfuerzos, han ido destinados a la comprensión del uso de los modelos para realizar predicciones y a la generación de sugerencias. Todo ello, dificultado por la masiva cantidad de información con la que trata el sistema, por lo que sería interesante realizar mejoras en estos apartados, como se explicará en el apartado de “6.2 Trabajo futuro de esta memoria”.

Según lo expuesto en este trabajo, es crucial el desarrollo de herramientas para realizar un seguimiento y dar apoyo a estudiantes de cursos online para conseguir reducir las altas tasas de abandono y de fracaso.

## 6.2 Trabajo futuro

La aplicación desarrollada, cuenta con la característica de ser escalable, permitiendo así nuevos desarrollos o mejoras de la misma. Se pondrá en conocimiento de la misma a los responsables de los MOOCs de la UAM, que podrían ofrecer la utilización de sus resultados en tiempo real con nuevas ediciones de los cursos.

Se podrían realizar ampliaciones para añadir nuevos indicadores a la plataforma. Por ejemplo, algunos que midan la eficiencia, el progreso, el tiempo, etc. Estos indicadores pueden requerir nuevos desarrollos si necesitan un post-procesado para su uso, por ejemplo, para medir progresos en la realización de problemas se requerirá conocer el total de problemas que contiene el curso. También será necesario desarrollar los extractores de estos nuevos indicadores para poder ser almacenados en la plataforma.

Debido a la gran cantidad de información almacenada en el sistema, sería interesante abandonar el uso de una base de datos relacional y optar por una base de datos NoSQL.

Sería posible realizar ampliaciones con nuevos métodos y algoritmos para la generación de modelos predictivos.

Como ya se ha comentado en este documento, el módulo de procesamiento tiene una gran carga computacional, por lo que sería interesante plantear mejoras de rendimiento y optimización para reducir los tiempos de cálculo.

En cuanto a las sugerencias, se pueden realizar mejoras en la clasificación de los usuarios, mejorando el sistema actual o incluso añadiendo más categorías. Como ya se ha comentado es posible en este documento también modificar los textos de las sugerencias o añadir nuevos.

El sistema debería en el futuro permitir exportar más información, generar más gráficas para que puedan ser utilizadas por los investigadores para realizar análisis más profundos sobre los usuarios, aportar más feedback para que el usuario sepa en todo momento que procesos se están ejecutando en el sistema, etc.





# Referencias

---

- [1] **Wikipedia** (Consultado en abril 2018). Macrodatos. Lugar de publicación: *Wikipedia website*. <https://es.wikipedia.org/wiki/Macrodatos>
- [2] **edX Inc.** (Consultado en abril 2018). About-us. Lugar de publicación: *edX website*. <https://www.edx.org/es/about-us>
- [3] **Wikipedia.** (Consultado en marzo 2018). Data anonymization. Lugar de publicación: *Wikipedia website*. [https://en.wikipedia.org/wiki/Data\\_anonymization](https://en.wikipedia.org/wiki/Data_anonymization)
- [4] **Wikipedia.** (Consultado en abril 2018). Educación a distancia. Lugar de publicación: *Wikipedia website*. [https://es.wikipedia.org/wiki/Educaci%C3%B3n\\_a\\_distancia](https://es.wikipedia.org/wiki/Educaci%C3%B3n_a_distancia)
- [5] **Online Learning Consortium.** (Consultado en mayo 2018). The Distance Education Enrollment Report 2017. Lugar de publicación: *Online Learning Consortium website*. [https://onlinelearningconsortium.org/news\\_item/new-study-six-million-students-now-enrolled-distance-education/](https://onlinelearningconsortium.org/news_item/new-study-six-million-students-now-enrolled-distance-education/)
- [6] **Wikipedia.** (Consultado en abril 2018). Massive Open Online Course. Lugar de publicación: *Wikipedia website*. [https://es.wikipedia.org/wiki/Massive\\_Open\\_Online\\_Course](https://es.wikipedia.org/wiki/Massive_Open_Online_Course)
- [7] **Universidad Internacional de Valencia.** (Consultado en marzo 2018). Características MOOC. Lugar de publicación: *Universidad Internacional de Valencia*. <https://www.universidadviu.es/caracteristicas-que-debe-cumplir-un-curso-mooc-en-espanol/>
- [8] **Atiaja Atiaja, Lourdes.** (Consultado en marzo 2018). Los MOOCs: evolución, problemas y perspectivas hasta el 2015. Lugar de publicación: *Universidad Internacional de Valencia*. <https://www.universidadviu.es/caracteristicas-que-debe-cumplir-un-curso-mooc-en-espanol/>
- [9] **Macías, V.** (Consultado en noviembre 2017). Trabajo de Fin de Grado: edX-MAS. Lugar de publicación: *Universidad Autónoma de Madrid (Escuela Politécnica Superior)*.
- [10] **Camarena, L.** (Consultado en febrero 2018). Trabajo de Fin de Grado: edX-MAS+. Lugar de publicación: *Universidad Autónoma de Madrid (Escuela Politécnica Superior)*.
- [11] **Wikipedia.** (Consultado en noviembre 2017). R (Lenguaje de programación). Lugar de publicación: *Wikipedia website*. [https://es.wikipedia.org/wiki/R\\_\(lenguaje\\_de\\_programaci%C3%B3n\)](https://es.wikipedia.org/wiki/R_(lenguaje_de_programaci%C3%B3n))
- [12] **Wikipedia.** (Consultado en noviembre 2017). Python (Lenguaje de programación). Lugar de publicación: *Wikipedia website*. <https://es.wikipedia.org/wiki/Python>
- [13] **R Users.** (Consultado en noviembre 2017). Caret, Classification and Regression Training. Lugar de publicación: *Cran Project website*. <https://cran.r-project.org/web/packages/caret/index.html>
- [14] **R Users.** (Consultado en octubre 2017). DataTables. Lugar de publicación: *Cran Project website*. <https://cran.r-project.org/web/packages/DT/index.html>
- [15] **R Users.** (Consultado en octubre 2017). Ggplot2. Lugar de publicación: *Cran Project website*. <https://cran.r-project.org/web/packages/ggplot2/index.html>
- [16] **R Users.** (Consultado en noviembre 2017). Shiny. Lugar de publicación: *Cran Project website*. <https://cran.r-project.org/web/packages/shiny/index.html>
- [17] **R Users.** (Consultado en abril 2018). Shinydashboard. Lugar de publicación: *Cran Project website*. <https://cran.r-project.org/web/packages/shinydashboard/index.html>

- [18] Python Software Foundation. (Marzo 2018). Python Software Foundation. Lugar de publicación: *Python Software Foundation website*. <https://www.python.org/psf/>
- [19] **Wikipedia**. (Consultado en noviembre 2017). PostgreSQL. Lugar de publicación: *Wikipedia website*. <https://es.wikipedia.org/wiki/PostgreSQL>
- [20] **Ayse Saliha Sunar, Su White, Nor Aniza Abdullah, and Hugh C. Davis**. (Consultado en junio 2018). How learners' interactions sustain engagement: a MOOC case study. Lugar de publicación: *TAL Journal*
- [21] **C. O. Rodriguez**. (Consultado en junio 2018). MOOCs and the AI-Stanford like courses: Two successful and distinct course formats for massive open online courses. Lugar de publicación: *European Journal of Open, Distance and E-Learning*, vol. 15.
- [22] **M.A. Chatti, A.L. Dyckhoff, U. Schroeder, and H. Thüs**. (Consultado en junio 2018). A Reference Model for Learning Analytics. Lugar de publicación: *International Journal of Technology Enhanced Learning (IJTEL)*
- [23] **Alexander Walker y Luca Braglia**. (Consultado en mayo 2018). Openxlsx. Lugar de publicación: *Cran Project website*. <https://cran.r-project.org/web/packages/openxlsx/index.html>
- [24] **D.F.O., Onah, J. Sinclair, R. Boyatt**. (Consultado en mayo 2018). Dropout Rates of Massive Open Online Courses: Behavioural Patterns. Lugar de publicación: *The University of Warwick (United Kingdom)*
- [25] **R. Cobos, A. Wilde, Zaluska**. (Consultado en mayo 2018). Predicting attrition from Massive Open Online Courses in FutureLearn and edX. comparing attrition prediction in FutureLearn and edX MOOCs. Lugar de publicación: *In Proc of LAK FutureLearn Workshop in the Learning Analytics and Knowledge 2017 Conference (LAK 17), Canadá, 13-17 Marzo 2017*
- [26] **J.A. Ruipérez-Valiente, R. Cobos, P.J. Muñoz-Merino, A. Andújar, C. Delgado-Kloos**. (Consultado en mayo 2018). Early Prediction and Variable Importance of Certificate Accomplishment. Lugar de publicación: *In Proc of the European MOOC Stakeholder Summit 2017 (eMOOCs 2017)*. Leganés, Madrid, España, 22-26 Mayo 2017
- [27] **L. Olmos**. (Consultado en mayo 2018). Sistema informático para el análisis de datos en entornos educativos (Computer system for data analysis in educational environments). Lugar de publicación: *Universidad Autónoma de Madrid*. España. Bachelor dissertation. 2018.



## Glosario

---

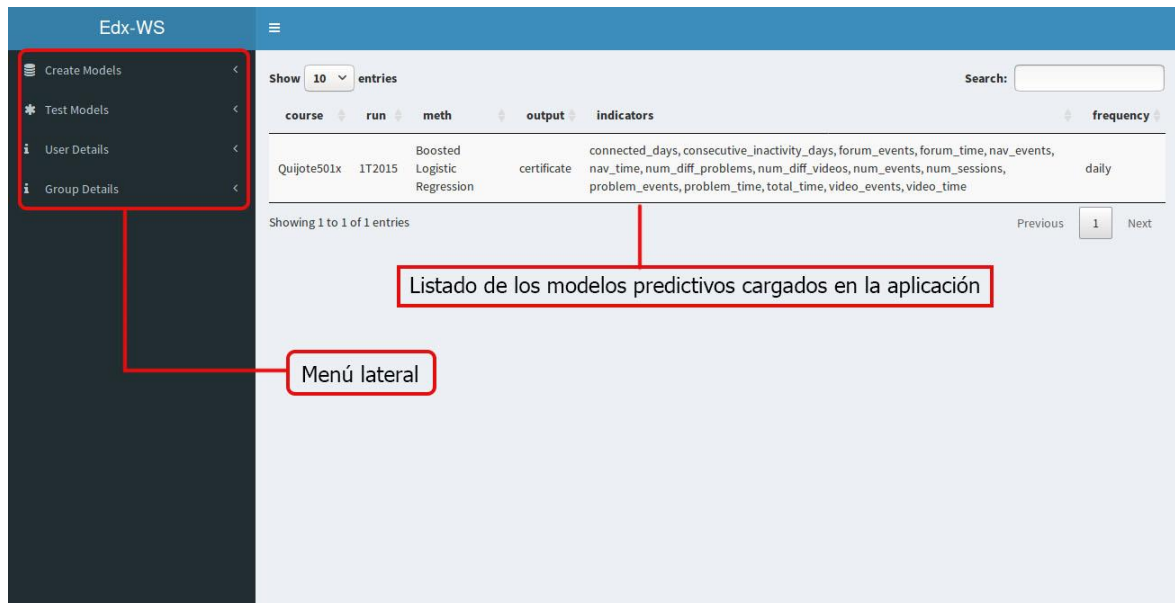
MOOC	Massive Open Online Courses
COMA	Cursos Online Masivos Abiertos
edX	Plataforma de cursos MOOC fundada por la Universidad de Harvard y el MIT
LAK	Learning Analytics and Knowledge
L@S	ACM Conference on Learning @ Scale
eMOOC	European MOOC Stakeholder Summit
EDUCON	Global Engineering Education Conference
UAM	Universidad Autónoma de Madrid
UAMx	Oficina para cursos MOOC de la UAM
edX-MAS	Model Analyzer System for edX MOOC. Desarrollado por Víctor Macías.
edX-MAS+	Model Analyzer System for edX MOOC, versión extendida. Desarrollado por Lara Olmos.
edX-WS	Warning System for edX MOOC



## Anexos

### A Manual de usuario

Pantalla inicial de la aplicación (Por defecto es la pantalla de “Create Models”). Se compone de un menú lateral, que nos permite navegar por las diferentes pantallas de la aplicación y de una parte central que nos permite visualizar todos los datos. (Véase Tabla 3-1: RF-01)



**Figura A-1: Explicación pantalla de inicio**

En la pantalla de “Create Models”, podemos observar en la parte central el listado de modelos predictivos ya cargados en el sistema. Este listado está representado en una tabla, la cual nos permite elegir el número de filas que visualizaremos por página y buscar por cualquier campo entre todos los resultados.

En el menú lateral de esta pantalla se encuentra el botón (“Show Available Models”), que nos permite mostrar en la parte central el listado de modelos. (Véase Tabla 3-1: RF-01)

The screenshot shows the 'Create Models' interface in Edx-WS. The left sidebar contains configuration options: 'Course' (TxEtj201x), 'Course editions' (1T2015 selected), 'Output' (certificate), 'Meth' (Boosted Logistic Regression), and 'Indicators' (num\_events, num\_sessions, total\_time, nav\_events). The main area displays a table of loaded models. Annotations include:

- Show 10 entries**: Points to the table's pagination controls.
- Search:**: Points to the search input field.
- Table headers**: course, run, meth, output, indicators, frequency.
- Table data**:
 

course	run	meth	output	indicators	frequency
Quijote501x	1T2015	Boosted Logistic Regression	certificate	connected_days, consecutive_inactivity_days, forum_events, forum_time, nav_events, nav_time, num_diff_problems, num_diff_videos, num_events, num_sessions, problem_events, problem_time, total_time, video_events, video_time	daily
- Número de registros que se muestran simultáneamente en la tabla**: Points to the 'Show 10 entries' text.
- Listado de modelos cargados**: Points to the table itself.
- Búsqueda por cualquier campo de la tabla**: Points to the search input.
- Paginación**: Points to the 'Previous', '1', and 'Next' buttons.
- Botón para visualizar la pantalla de creación y listado de modelos**: Points to the 'Show Available Models' button in the sidebar.

**Figura A-2: Explicación pantalla de modelos predictivos cargados**

En el menú lateral de esta pantalla podremos elegir los parámetros necesarios (curso, edición, output, algoritmo, indicadores, etc.) para generar un modelo predictivo nuevo. (Véase Tabla 3-1: RF-01)

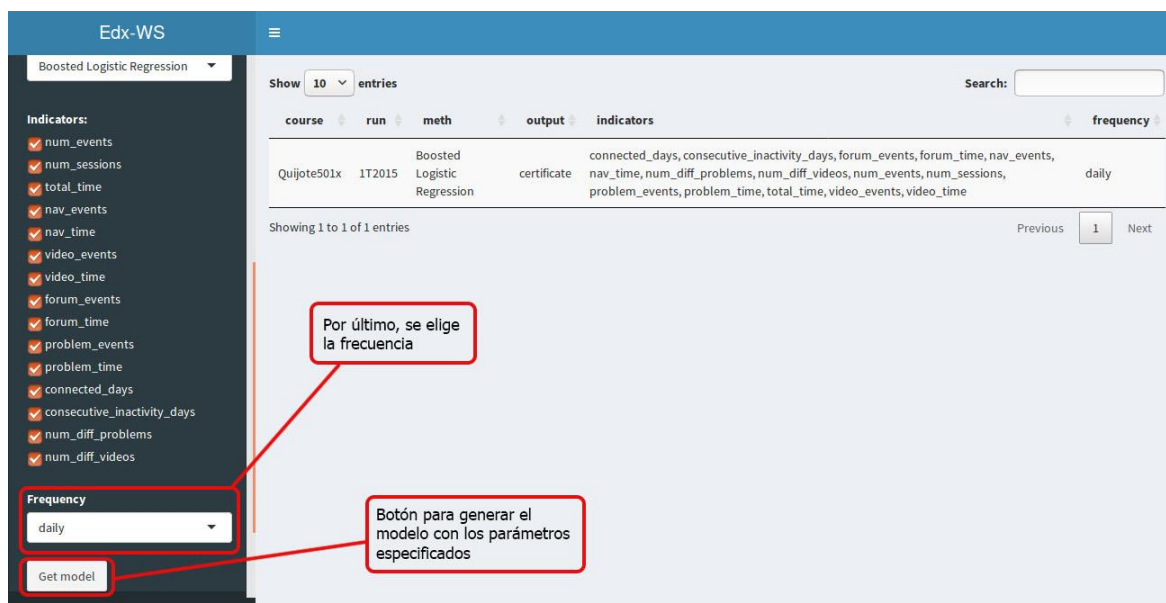
The screenshot shows the same 'Create Models' interface. An annotation points to the sidebar configuration options:

- Parámetros de configuración para la generación del modelo predictivo. Curso, edición, output, algoritmo, indicadores..**: Points to the sidebar containing Course, Course editions, Output, Meth, and Indicators.

**Figura A-3: Explicación de parámetros para generación de modelos**

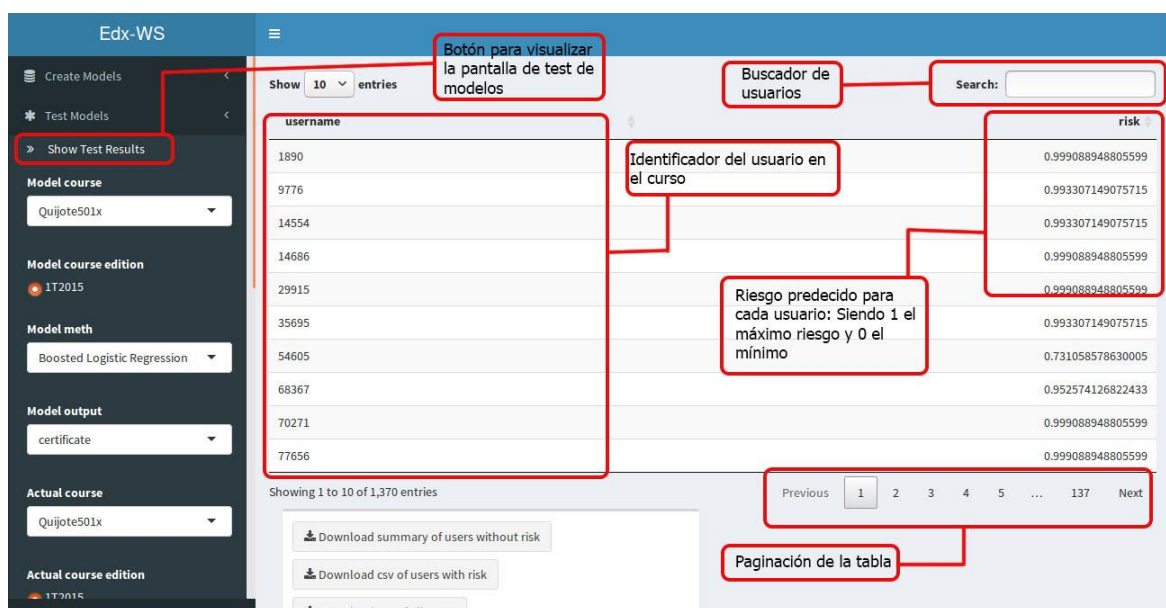


Una vez rellenos todos los parámetros, incluida la frecuencia, podemos generar un modelo predictivo nuevo presionando el botón “Get model”. (Véase Tabla 3-1: RF-01)



**Figura A-4: Explicación de parámetros para generación de modelos**

En pestaña “Test Models” del menú lateral, se encuentra el botón “Show Test Results”, que nos permite visualizar los resultados del test de un modelo predictivo. La tabla resultante, muestra un listado de los usuarios en riesgo de no obtener certificado o de no finalizar el curso. En esta tabla, podremos ver el riesgo asociado a cada usuario. Permitiendo la paginación y la búsqueda por “username”. (Véase Tabla 3-2: RF-02)



**Figura A-5: Explicación pantalla de visualización de los resultados del test**

En el menú lateral de esta pantalla, podremos seleccionar los parámetros necesarios para realizar un nuevo test. Para ello, primero se seleccionará un curso, una edición, un algoritmo y un output que identifica de manera única un modelo predictivo ya cargado en el sistema. (Véase Tabla 3-2: RF-02)

The screenshot shows the Edx-WS interface. On the left sidebar, under 'Show Test Results', there are four dropdown menus: 'Model course' (set to 'Quijote501x'), 'Model course edition' (set to '1T2015'), 'Model meth' (set to 'Boosted Logistic Regression'), and 'Model output' (set to 'certificate'). These are highlighted with a red box. A red callout box points to the 'Model course edition' dropdown with the text: 'Selección del modelo con el que se realizará el test. Curso, edición, algoritmo y output'. The main area displays a table with columns 'username' and 'risk'. The table shows 10 entries, with the first entry having a risk of 0.999088948805599. Below the table, there are download buttons for 'summary of users without risk', 'csv of users with risk', and 'csv of all users'.

**Figura A-6: Explicación de la selección del modelo para realizar el test**

Una vez seleccionado el modelo que se usará para realizar test, seleccionaremos los nuevos datos con los que realizar el test. Podremos escoger el curso, la edición, los indicadores, etc. (Véase Tabla 3-2: RF-02)

The screenshot shows the Edx-WS interface. On the left sidebar, under 'Actual course', there is a dropdown menu set to 'Quijote501x'. Below it, under 'Actual course edition', there are three radio buttons: '1T2015' (selected), '3T2016', and '3T2015'. These are highlighted with a red box. A red callout box points to the 'Actual course edition' radio buttons with the text: 'Selección del curso del cual se obtendrán los datos para realizar el test. Curso, edición, indicadores...'. Below the radio buttons, there is a section 'Indicators:' with a list of checkboxes, all of which are checked: 'num\_events', 'num\_sessions', 'total\_time', 'nav\_events', 'nav\_time', 'video\_events', 'video\_time', 'forum\_events', 'forum\_time', 'problem\_events', 'problem\_time', 'connected\_days', 'consecutive\_inactivity\_days', 'num\_diff\_problems', and 'num\_diff\_videos'. The main area displays a table with columns 'username' and 'risk'. The table shows 10 entries, with the first entry having a risk of 0.999088948805599. Below the table, there are download buttons for 'summary of users without risk', 'csv of users with risk', and 'csv of all users'.

**Figura A-7: Explicación de la selección del curso para realizar el test**

Para finalizar, y después de seleccionar hasta que día se realizará el test, se podrá presionar el botón “Test”, el cual lanzará en el servidor un proceso en que se evaluará para cada día el riesgo de cada usuario en base a sus indicadores. (Véase Tabla 3-2: RF-02)

The screenshot shows the Edx-WS interface. On the left, there is a sidebar with a list of indicators and a 'Select day' slider. The main area displays a table of user risk data. Red boxes and arrows highlight the 'Test' button, the 'Select day' slider, and the table rows.

**Indicators:**

- num\_events
- num\_sessions
- total\_time
- nav\_events
- nav\_time
- video\_events
- video\_time
- forum\_events
- forum\_time
- problem\_events
- problem\_time
- connected\_days
- consecutive\_inactivity\_days
- num\_diff\_problems
- num\_diff\_videos

**Select day:** 1 to 56

**Test**

**Table:**

username	risk
1890	0.999088948805599
9776	0.993307149075715
14554	0.993307149075715
14686	0.999088948805599
29915	0.999088948805599
35695	0.993307149075715
54605	0.731058578630005
68367	0.952574126822433
70271	0.999088948805599
77656	0.999088948805599

Showing 1 to 10 of 1,370 entries

Previous 1 2 3 4 5 ... 137 Next

Download summary of users without risk

Download csv of users with risk

Download csv of all users

**Figura A-8: Explicación de la selección del curso para realizar el test**

Una vez obtenidos los resultados del test, podremos exportar diferentes datos a ficheros csv (usuarios con riesgo, usuarios sin riesgo, estadísticas, sugerencias). También podremos seleccionar un usuario de la tabla para ver sus detalles en la pestaña “User Details” (Véase Figura A-11).

The screenshot shows the Edx-WS interface. On the left, there is a sidebar with a list of indicators and a 'Select day' slider. The main area displays a table of user risk data. Red boxes and arrows highlight the 'User Details' tab, the table rows, and the export buttons.

**Indicators:**

- num\_events
- num\_sessions
- total\_time
- nav\_events
- nav\_time
- video\_events
- video\_time
- forum\_events
- forum\_time
- problem\_events
- problem\_time
- connected\_days
- consecutive\_inactivity\_days
- num\_diff\_problems
- num\_diff\_videos

**Select day:** 1 to 56

**Test**

**User Details**

**Table:**

username	risk
1890	0.999088948805599
9776	0.993307149075715
14554	0.993307149075715
14686	0.999088948805599
29915	0.999088948805599
35695	0.993307149075715
54605	0.731058578630005
68367	0.952574126822433
70271	0.999088948805599
77656	0.999088948805599

Showing 1 to 10 of 1,370 entries

Previous 1 2 3 4 5 ... 137 Next

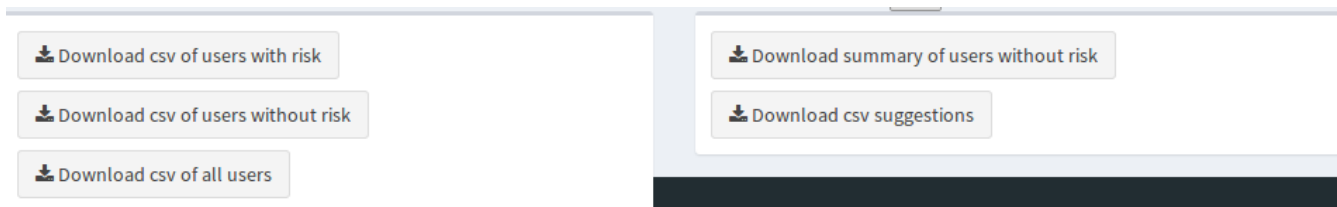
Download summary of users without risk

Download csv of users with risk

Download csv of all users

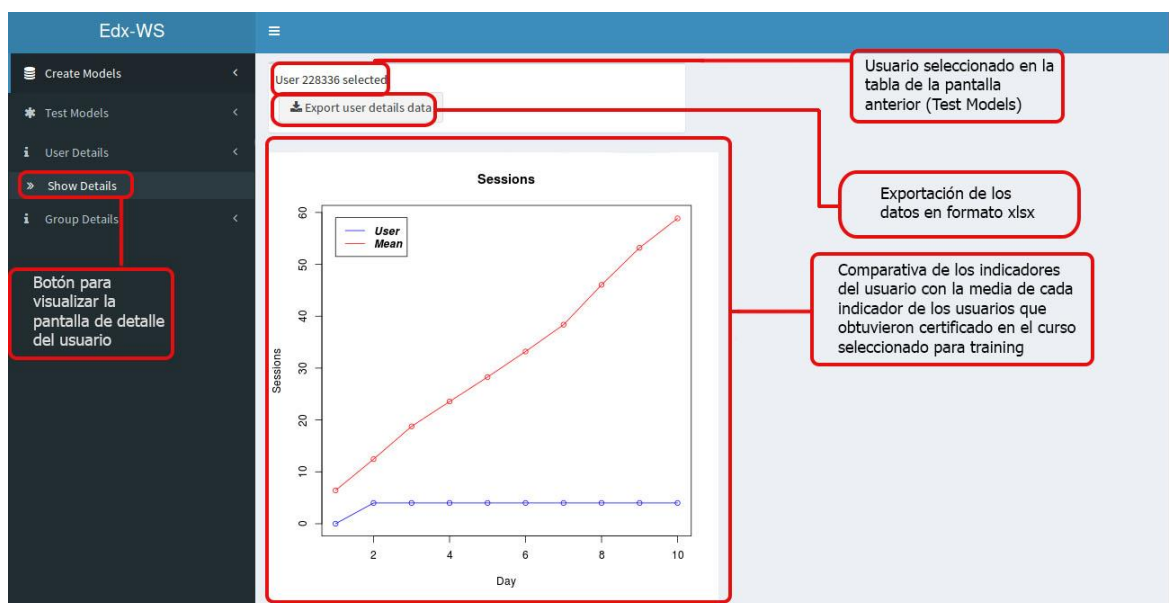
**Figura A-9: Explicación para consultar detalle y exportaciones**

Listado de botones que nos permiten exportar a *csv* los datos obtenidos del test y generar un fichero con las sugerencias para los usuarios en riesgo. (Véase Tabla 3-3: RF-03 y Tabla 3-4: RF-04)



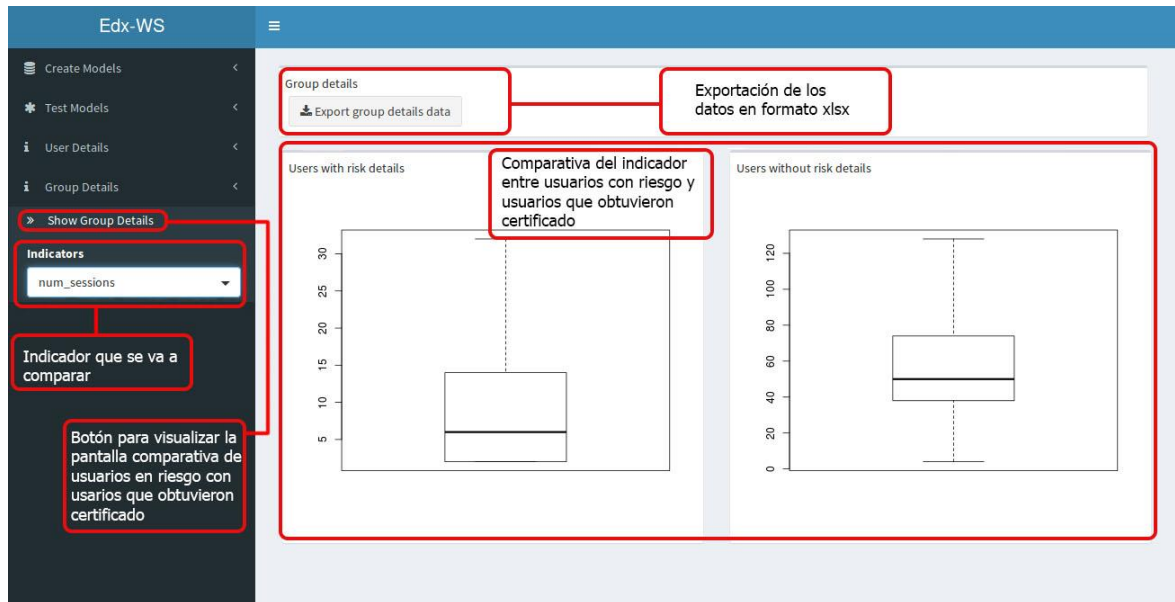
**Figura A-10: Descarga de ficheros (Exportaciones y sugerencias)**

En la pantalla “User Details”, podremos ver las diferentes gráficas de cada indicador para el usuario seleccionado previamente en la pestaña de “Test Results” (Ver Figura A-9). Estas gráficas son una comparativa entre el usuario seleccionado y los valores medios de cada indicador para los usuarios que obtuvieron certificado en años anteriores. También encontramos en la parte superior el botón “Export user details data”, el cual genera un fichero *xlsx* con los datos utilizados para generar las gráficas. (Véase Tabla 3-5: RF-05)



**Figura A-11: Explicación pantalla de visualización de detalle de un usuario**

En la pantalla “Group Details”, podemos observar una comparativa con gráficas Boxplot de los valores de cada indicador (seleccionado en el menú lateral). En el lado izquierdo tendremos los valores de los indicadores de los usuarios en riesgo y en el lado derecho los valores de los indicadores para los usuarios sin riesgo. También encontramos en la parte superior el botón “Export group details data”, el cual genera un fichero *xlsx* con los datos utilizados para generar las gráficas. (Véase Tabla 3-6: RF-06)

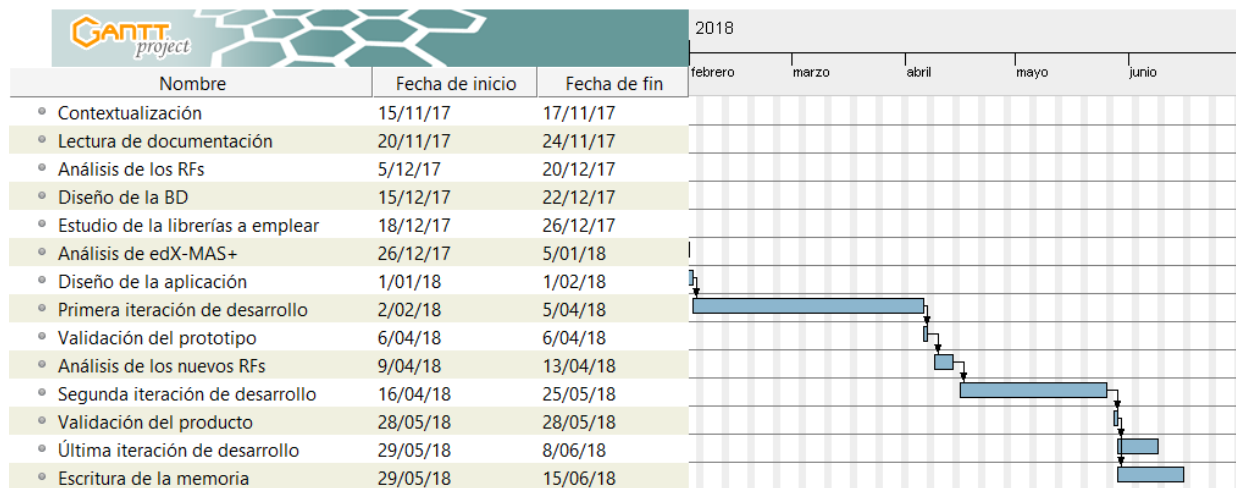


**Figura A-12: Explicación pantalla de visualización de detalle de grupos**

## B Diagrama de Gantt



**Figura B-1: Diagrama de Gantt 1**



**Figura B-2: Diagrama de Gantt 2**